BT

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 9/22, 15/55, C12P 19/34

A1

(11) International Publication Number: **WO 96/40882**

(43) International Publication Date: 19 December 1996 (19.12.96)

(21) International Application Number: PCT/US96/09315

(22) International Filing Date: 6 June 1996 (06.06.96)

(30) Priority Data:
08/484,174 7 June 1995 (07.06.95) US

(71) Applicant: THE OHIO STATE UNIVERSITY [US/US]; 1960 Kenny Road, Columbus, OH 43210-1063 (US).

(72) Inventors: TSAI, Ming-Daw; 2853 Pickwick Drive, Columbus, OH 43221 (US). HUANG, Baohua; 700 Ashtabula Court, Columbus, OH 43210 (US).

(74) Agent: GOLRICK, Mary, E.; Calfee, Halter & Griswold, Suite 1400, 800 Superior Avenue, Cleveland, OH 44114 (US).

(81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: ARTIFICIAL RESTRICTION ENDONUCLEASE

(57) **Abstract**

The present invention provides novel, artificial, restriction endonucleases comprising the DNA binding motif of the transcription factor Spl and the C-terminal DNA cleavage domain of FoKI restriction endonuclease. The new restriction endonucleases are designated herein as "splase enzyme" and "Hsplase enzyme". The restriction endonuclease of the present invention recognizes a 10 base nucleotide sequence and thus, since the probability of these particular 10 base sequences occurring frequently in a DNA sample is low, the restriction endonucleases only rarely cleave the DNA sample. Thus, the Splase enzyme and Hsplase enzyme are particularly useful cutting DNA into large fragments rather than into a myriad of fragments as is common with conventional restriction endonucleases. The invention also relates to artificial fusion genes which encode the artificial restriction endonucleases including the artificial restriction endonucleases Splase enzyme and Hsplase enzyme. The DNA molecule, that is the genes that encode the Splase enzyme and Hsplase enzyme are designated herein as the "Splase gene" and the "HSplase gene", respectively. The invention also relates to vectors and cells that contain the artificial genes.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

ARTIFICIAL RESTRICTION ENDONUCLEASE

BACKGROUND OF THE INVENTION

Restriction endonucleases are enzymes that cleave double stranded DNA at specific points. Restriction endonucleases have been isolated from a variety of organisms and employed as a valuable tool in recombinant DNA technology. Each restriction endonuclease recognizes a certain base sequence and only that sequence. Each restriction endonuclease is a tool permitting flexibility in the manipulation and assembly of DNA in-vitro; new restriction endonucleases are desirable since they increase the techniques for the in-vitro manipulation of DNA.

SUMMARY OF THE INVENTION

The present invention provides novel, artificial, restriction endonucleases comprising the DNA binding motif of the transcription factor Spl and the C-terminal DNA cleavage domain of FoKI restriction endonuclease. The new restriction endonucleases are designated herein as "splase enzyme" and "Hsplase enzyme". The restriction endonuclease of the present invention recognizes a 10 base nucleotide sequence and thus, since the probability of these particular 10 base sequences occurring frequently in a DNA sample is low, the restriction endonucleases only rarely cleave the DNA sample. Thus, the Splase enzyme and Hsplase enzyme are particularly useful for cutting DNA into large fragments rather than into a myriad of fragments as is common with conventional restriction endonucleases. The invention also relates to artificial fusion genes which encode the artificial resriction endonucleases including the artifical resrtiction endonucleases Splase enzyme and Hsplase enzyme. The, DNA molecule, that is the genes that encode the Splase enzyme and Hsplase enzyme are designated herein as the "Splase gene" and the "HSplase gene," respectively. The invention also relates to vectors and cells that contain the artificial genes.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a mobility-shift gel in which: DNA substrate Seq. ID No. 16 incubated with Splase was applied to Lane 1. The DNA substrate Seq. ID No. 16 was applied to Lane 2. The DNA substrate Seq. ID No. 16

incubated with splase enzyme in the presence of the competitor DNA, which lacked an spl site, was applied to lane 3. The DNA substrate Seq. ID No. 16 and competitor DNA was applied to lane 4. The competitor DNA incubated with splase enzyme was applied to lane 5. The competitor DNA was applied to lane 6. The gel which was subjected to electrophoresis in TAE buffer without EDTA, demonstrates the ability of the HSplase enzyme to specifically bind to the Spl site;

Figure 2A is a gel containing the following: the pUC19 DNA digested by Hsplase was applied to lane 1; pUC19 DNA without Hsplase was applied to lane 2; DNA molecular weight marker VII from Boehringer Mannheim was applied to lanes 3; the pUC3spl digested by H-splase was applied to lane 4; the pUC5SPl DNA was applied to lane 5; the pUC5SPl incubated with Hsplase was applied to lane 6; pUC3Spl was applied to lane 7; and spUC5Spl digested by Hsplase was applied to lane 8. The gel was then subject to electrophoresis; the results demonstrate HSplase cleaved the closed circular DNA with 5 Spl sites into linear form;

Figure 2B is a gel in which: pUC19 substrate after double digestion by Hsplase and AWN was applied to lane 1; pUC5SPl substrate after double digestion was applied to lane 2; and the DNA molecular weight marker VII from Boehringger Mannheim was applied to lane 3. The gel was then subjected to electrophoresis; the gel confirms that cleavage of the closed circular DNA is specific and near Spl sites; and

Figure 3 is a gel in which: molecular weight markers were applied to lanes 1 and 8; AlWNI linearized pUC19 was applied to lane 2; AlWNI linearized pUC190 incubated with Hsplase was applied to lane 3; AlWNI linearized pUC5Spl was applied to lane 4; AlWNI linearized pUC5Spl incubated with HSplase was applied to lane 5; BamHI linearized pUC-BENN-CAT was applied to lane 6. BamHI linearized pUC-BENN-CAT incubated with HSplase was applied to lane 7. The gels were then subject to electrophoresis. The gels demonstrate that HSplase cleaved linear DNA specifically near Spl sites.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides artificial chimeric proteins, specifically enzymes, designated "splase" enzyme and "Hsplase" enzyme, each of said enzymes comprising the DNA binding motif of the transcription factor Spl and the C-terminal DNA cleavage domain of FokI restriction endonuclease. The invention also relates to fusion genes encoding the chimeric enzymes splase and Hsplase.

The gene encoding the splase enzyme is shown in SEQ ID NO 1 and the amino acid sequence of the splase enzyme is shown in SEQ ID NO 2. The gene encoding the Hsplase enzyme is shown in SEQ ID NO 3 and the amino acid sequence of the splase enzyme is shown in SEQ ID NO 4.

The Splase Enzyme

The splase enzyme comprises the amino acid sequence as shown in Sequence ID No. 2. The Hsplase enzyme comprises the amino acid sequence of splase, and also comprises histidine residues, preferably 6 histidines at the N-terminus. The Hsplase enzyme comprises the amino acid sequence as shown in Sequence ID No. 4. The first two residues at the N-terminus of Splase enzyme are methionine and valine. The next sequence of 92 amino acids at the N-terminus includes the 92 amino acid zinc finger motif of Spl except that the residue at position 92 of the motif is valine rather than glycine. This change was made for the convenience of gene fusion. The 203 amino acid sequence at the C-terminus of the splase enzyme is preferably identical to the 203 amino acids of the C-terminal cleavage domain of FokI.

The zinc-finger domain of Spl, a 92-amino acid peptide sequence encoded by the zinc-finger motif of the transcription factor Spl is a sequence-specific DNA-binding domain. This domain recognizes several closely related 10 base pair Spl DNA binding sites; recognized spl binding sites include:

Sequence ID No. 5

5'-G(T)GG GCG GG(A)G(A)C(T)-3'.

The Hsplase enzyme further recognizes and bind to the following Spl sites:

5'-GGGGCGGGGC-3' Sequence ID No. 6

5'-GAAGCGTGGC-3' Sequence ID No. 7

5'-TGGGCGGGAC-3' Sequence ID No. 8

5'-GGGGAGTGGC-3' Sequence ID No. 9

Thus an example of the splase sites includes:

Sequence ID No. 10

5'-G(T)G(A)G(A) GC(A)G G(T)G(A)G(A)C(T)-3'.

The splase enzyme cleaves linear DNA and circular DNA. The splase enzyme cleaves near the Spl site.

Gene Construction

The splase gene was constructed and is shown in Seq ID NO 1 and the Hsplase gene was constructed and is shown in Seq ID NO 3. Polymerase chain reaction techniques were used to amplify and subclone the DNA fragment encoding the Spl zinc finger domain. The template used was plasmid pSpl-516C. The pSpl-516C contains DNA Sequence encoding the C-terminal 516 amino acids of Spl. The oligonucleotide primer set used is shown below:

5' primer :

5'-g tcc atg gct aaa aag aaa cag cat att tgc cac-3' Seq.ID.No. 11

3' primer :

3'-c tgg gtg gtc tta ttc ttc cat ggc c-5' Seq.ID.No. 12

First, 100 ng of the plasmid pSpl-516C was amplified in 100 µl reaction volume, which contained: 10mM Tris-HCl at pH 8.8; 10 mM KCl; 1 mM $MgCl_2$; 2 µM of each of the primers represented above by Sequence ID Nos 11 and 12; 10 µm each of dATP, dTTP, dGTP and dCTP, 0.003% Tween 20 v/v; and 3 units of ULTIma DNA polymerase from Perkin-Elmer. The reaction was "hot started" using Ampliwax beads from Perkin Elmer. The PCR was conducted in a Perkin-Elmer DNA Thermal Cycler employing the following cycling parameters: an initial denaturation of 2 minutes at 95° C was followed by 30 cycles of amplification for 2 minutes at 95° C, 2 minutes at 60° C, 3 minutes at 72° C.

The reaction was run in duplicate and the PCR products were combined and extracted with phenol/chloroform (1:1) and DNA was precipitated with ethanol and dissolved in 20 µl of TE which contains 10 mM Tris-HCl, pH 8.0, and 1 mM EDTA. The PCR product is a DNA fragment coding for the zinc finger domain of Spl. To prepare the PCR product for ligation, 5 µl of the DNA was digested with NcoI/KpnI from Boehringer Mannheim, Germany according to the manufacturer's instructions. The digested DNA was separated by electrophoresis on 0.8% agarose gel in TAE which contained 0.4 M Tris-HCl, 0.013 M sodium acetate, and 0.2 mM EDTA, at pH 8.0. The DNA band of approximately 320 base pairs was cut out of the gel, purified using glassmilk as suggested by manufacturer, BI0101, Inc., Vista, CA, and dissolved in 10 µl of the ddH_2O. 2 µl of DNA was then ligated into 0.1 µg of a NcoI/KpnI cleaved pTrc99A DNA at 16°C for 4 hours. The ligation mixture was transformed into JM105 competent cells and positive clones were identified. The resulting plasmid was designated "pTRCSpl". Detailed protocol is described in the following:

To clone the gene fragment coding Spl zinc finger domain amplified by PCR and subsequently fused it with the PCR fragment coding the FokI nuclease domain, E. coli JM105 was used. The cells were made competent by the protocol described below and transformed with the DNA ligation mixture. Plasmid DNAs were then isolated from individual colonies and positive colonies containing vector with gene fragment insert were identified by restriction mapping. Preparation of JM105 Competent Cells

JM105 was streaked onto the surface of an M9 agar plate and incubated 16 hours at 37°C. To 4 ml of LB medium (10 g yeast extract, 5 g tryptone and 10 g NaCl per liter), a single colony from the plate was inoculated. The culture was incubated 8 hours at 37°C with shaking. 1 ml of this culture was used to inoculate 200 ml SOB media (20 g trypton, 5 g yeast extract and 0.5 g NaCl) supplemented with 2 ml

each of 1 M $MgSO_4$ and 1 M $MgCl_2$. The culture was then incubated at 37°C with shaking for 2-3 hours until absorbance at 600 nm reached 0.3 Then 100 ml aliquot of the culture was cooled on ice for 10 min. Cells were then centrifuged at 6,000 X g for 5 minutes. The supernatant was discarded and cells were resuspended in 8 ml of ice-cold TFBI (0.59 g KOAC, 1.49 g KCl, 0.29 g $CaCl_2 \cdot H_2O$, 1.98 g $MnCl_2 \cdot H_2O$, and 30 g glycerol per 200 ml, pH 5.8) and left on ice for 10 min. The cells were centrifuged again at 6,000 X g for 5 minutes. Tthe supernatant was discarded and cells were resuspended in 8 ml of ice-cold TFBII (0.42 g MOPS, 0.15 g KCl, 2.21 g $CaCl_2 \cdot 2\ H_2O$ and 30 g glycerol per 200 ml, pH 5.8). The cell suspension (now competent) was then frozen in 200 µl aliquots in liquid nitrogen and stored at -70°C.

Transformation of JM105 Competent Cells

200 µl of thawed JM105 cells were incubated with 5 µl of DNA ligation mixture on ice for 45 minutes. The cells were then heat shocked at 42°C for 90 seconds and placed on ice for 5 minutes. 800 µl of LB media was added to the transformation mixture and incubated at 37°C for one hour. 200 µl of the transformation mixture was then spread onto an LB agar plate containing 100 µg/ml of ampicillin. The plate was incubated at 37°C overnight and colonies were then visible.

To screen for positive colonies, the following miniprep of plasmid DNA was conducted. Individual colonies were inoculated into 4 ml LB media containing 100 µg/ml ampicillin and cultured at 37°C with shaking for 10 hours. 1.5 ml of the culture was centrifuged in a microcentrifuge for 1 minute and the supernatant was decanted, leaving 100 µl with the pellet. The pellet was then resuspended in the remaining supernatant by vortexing for 5 second 300 µl of TENS buffer (10 mM Tris-HCl, pH 8.0, 0.1 M NaOH, 1 mM EDTA and 0.5% SDS) was added, the tube was vortexed for 5 seconds, 150 µl of 3 M NaOAC, pH 5.2 was then added and the tube was vortexed again. The sample was then centrifuged in a microcentrifuge for 2 minute and the supernatant was transferred to a fresh tube and mixed with two volumes of cold ethanol. DNA was brought down by centrifuging in a microcentrifuge 2 minute After washing twice with 70% ethanol, DNA was dried in a SpeedVac and dissolved in 50 µl TE. The solution was spun again and supernatant containing DNA was transferred to a new tube and stored at 4°C.

Screening Positive Colonies by Restriction Mapping

To identify clones with gene insert, plasmid DNA from miniprep was digested with the same restriction enzyme set from Boehringer Mannheim Company, Germany, which was used for cloning (in this case, they are NcoI and KpnI) according to the manufacturer's instructions. The mixture was then subjected to electrophoresis on a 0.7% agarose gel, stained with ethidium bromide, and the DNA fragments were

visualized under UV light. Positive clones with gene insert give a restriction fragment of the right size for the gene insert. In this case, the NcoI/KpnI fragment coding Spl zinc finger domain was 320 base pairs.

The DNA fragment encoding the C-terminal DNA cleavage domain of FokI was also amplified and cloned employing polymerase chain reaction techniques. The template was genomic DNA isolated from *Flavobacterium okeanokoites* and the primer set used is shown below:

5' primer:

5'-cgg gta cct aat cgt ggt gtg act aag-3' Seq.ID. NO. 13

3' primer:

3'-tta ttg ccg ctc tat ttg aaa att cct agg cg-5' Seq.ID.NO. 14

The PCR conditions were similar to those used for amplification of the Spl coding sequence, except that 300 ng of the genomic DNA was used as the template and the annealing temperature was 55°C. The genomic DNA was isolated using the miniprep protocol as described as follows. Flavobacterium Okeanokoites was grown at 37°C for 48 hours with shaking in 3 ml medium containing per liter 10 g of trypton, 5 g of yeast extract, 2 g of NaCl and 4.4 K_2HPO_4. The cells were harvested by centrifugation in a microcentrifuge. The cell pellet was resuspended in 284 µl TE buffer which contained 10 mM Tris-HCl and 1 mM EDTA, pH 8.0. Next, 15 µl of 10% SDS and 3 µl of protease K (20 mg/ml) were added and the sample was incubated at 37°C for one hour. The sample was thoroughly mixed first with 50 µl of 5 M NaCl and then with 40 µl of CTAB/NaCl solution which contained 10% hexadecyltrimethyl ammonium bromide in 0.7 M NaCl. The sample was incubated at 65°C for ten minutes. The sample was subsequently extracted with 0.4 ml of chloroform and spun for five minutes in a microcentrifuge. The aqueous phase was transferred to a new tube and extracted with equal volume of phenol/chloroform (v/v 50%). The aqueous phase was transferred to a fresh tube and DNA was precipitated 2.5 volumes of ethanol. The sample was washed twice with 70% ethanol and dried in a SpeedVac. The sample DNA was dissolved in 20 µl double distilled water and 1 µl TE was added. The DNA concentration was 0.3 mg/ml as determined by UV absorbance at 260 nm.

Next the PCR generated FokI cleavage domain gene fragment was digested with KpnI/GamHI and gel purified. The 625 base pair fragment was excised, purified and then the fragment was ligated into the KpnI/GamHI-cleaved vector pTrcSpl. The resulting plasmid was designated "pTrcSplase". The pTrcSplase fuses, in frame, the 92-amino acid DNA binding domain of Spl to the C-terminal 203 amino acid FokI cleavage domain. The entire Splase gene was subsequently cut from

pTrc-Splase with NcoI/BamHI and then subcloned into a NcoI/GamHI cleaved pTO-N vector. A description and the construction of the PTO-N vector is described in "A Novel Express Vector for High-Level Synthesis and Secretion of Foreign Proteins in E. Coli: Overproduction of Bovine Pancreatic Phospholipase A_2", T. Deng, et al., Gene 93, 229-234 (1990). The resulting vector which is designated "pTO-Splase" directs the synthesis of Splase enzyme with the OmpA signal peptide fused to its N-terminus. The OmpA signal peptide directs the synthesized Splase enzyme into periplasmic space, thus overcoming the potential toxicity of Splase enzyme to E. coli.

Addition of Codons for a Histidine Tag to the Splase Gene

To facilitate the purification of the recombinant Splase enzyme overexpressed in E. coli, codons for six consecutive histidines were added to the Splase gene to produce the gene designated as the "HSplase gene". The 6-histidine tag at the N-terminus facilitates the purification of the Hsplase enzyme by metal-chelating chromatography with Novagene's His-bind resin.

The Splase gene was PCR-amplified using the pTrcSplase plasmid as a template. The 3' primer that was employed is the same primer as shown in Seq. ID No. 14. The 5' HSplase primer is shown below:

Seq. ID. No. 15

5'- g tcc atg gct cat cac cat cac cat cac aaa aag aaa cag cat att tgc cac-3'

The PCR conditions were the same as that used to amplify DNA fragment encoding Spl Zinc finger domain. The PCR-generated HSplase gene fragment was then digested with NcoI/Kpn I and the 320 base pair DNA fragment was gel purified. This fragment was subsequently ligated to a NcoI/KpnI-cleaved pTrc99A. The resulting vector is designated as "pTrcHSpl".

The KpnI/BamHI fragment coding for the FoKI nuclease domain was cut out from pTrc-Slase and ligated into pTrc-HSpl cleaved by the same enzyme set. The resulting plasmid is designated as pTrcH-Splase. The sequence of the HSplase gene was confirmed by dideoxy-sequencing of pTrcH-Splase dsDNA according to the method of Sanger, F., et al. (1977) *Proc. Natl. Acad. Sci. USA* 74, 5463-5467.

Expression of Splase and HSplase Genes

Both the Splase and HSplase genes were expressed in pTrc99A and pTO-N vectors. The plasmids used were pTRCSplase, pTOSplase, pTRCH-Splase and pTOH-Splase.

Purification of the Hsplase Enzyme from Cytosol

E. Coli strain JM105 cells transformed as described above, with pTrc-HSplase were grown in 2.5 liters of medium which contained: 12g yeast extract, 19 g tryptone, 10mM $MgCl_2$ per liter, and 100 μg/ml of ampicillin. The cells were grown at 37°C to an OD_{600} of 0.6 units and cooled to 25°C. The cells were induced with 0.4 mM isopropyl β-D-thiogalactoside. The cells were allowed to grow overnight and harvested by centrifugation and then resuspended at 3 ml/g wet weight in 1 X binding buffer which contained 6 mM imidazole; 0.5 M NaCl; 20 mM Tris-HCl at pH 7.9, at 4°C. Next 18 grams of the cells were disrupted in ice using a Branson sonicator. Sonication lasted for 45 seconds and was repeated 3 times. The sonicated cells were centrifuged at 15,000 X g for 25 minutes at 4°C. The supernatant was transferred to a new tube and the pH was adjusted to 7.9. The HSplase enzyme was purified from the supernatant by metal-chelating chromatography using His-bind resin from Novagene. In brief, the sample was filtered through a 0.45 μm filter and then loaded onto a 5 ml His-bind column equilibrated with 1X binding buffer. The column was washed with 10 vol of binding buffer and 10 vol of wash buffer which contained 60 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl at pH 7.9. The column was then eluted with a 20 ml elution buffer containing 500 mM imidazole, 0.5 M NaCl, 20 mM Tris-HCl at pH 7.9. The fractions that contained the HSplase enzyme were pooled and desalted on a Sephadex G-25 column equilibrated with enzyme buffer which contained 20 mM Tris-phosphate, pH 7.7, 50 mM NaCl. DTT was added to the final sample to bring the sample to a final concentration of 5 mM and the enzyme sample was frozen at -70°C. The yield was about 5-10 mg.

The Hsplase enzyme has a molecular weight of 35 kDa and accounted for about 10-20% of total cellular enzyme. This molecular weight is in agreement with the calculated molecular weight of 35 kDa for HSplase. The Hsplase enzyme was also purified to greater than 90% purity using His-bind column under denaturing conditions.

Control cells which contained only the pTrc99A vector were also grown and analyzed as above. The control cells did not make the Hsplase enzyme.

The Hsplase enzyme was then evaluated as discussed hereinafter.

Purification and Refolding of Hsplase from Inclusion Body

JM105 cells transfected with plasmid pTrc-HSplase were grown in 4 liters of 2 X TY containing 100 μg of ampicillin per ml at 37°C to an optical density at 600 of 0.7 units. The cells were then induced 8 hours with 1mM isopropyl β-D-thiogalactoside. The cells were harvested by centrifugation and then resuspended at 2.8 ml/g wet weight in binding buffer which contains 6 mM imdazole, 0.5 M NaCl, 20 mM

Tris-HCl, at pH 7.9 and 8 M urea. Cells were disrupted on a Branson sonicator for 5 minutes on ice. The sonicated cells were centrifuged at 15,000 X g for 25 minutes at 4°C. The inclusion body in the pellet was washed once with 50 ml of 1X binding buffer and centrifuged again. The inclusion body pellet was solubilized in 100 ml of 1 X binding buffer, which contained: 6 mM imidazole; 0.5 M NaCl; 20 mM Tris-HCl; at pH 7.9 and 8 M urea for 1 hour. The sample was then centrifuged as above and the supernatant was mixed with 10 ml of His-bind resin for batch binding. Binding was performed in a 250 round bottom flask and the sample was mixed by rotating for 1 hour. The slurry was then poured into a column. The column was washed with 10 volume of binding buffer and 10 volume of wash buffer which contained: 20 mM imidazole; 0.5 M NaCl; 20 mM Tris-HCl at pH 7.9, and 8 M urea. The column was then eluted with a 60 ml elution buffer. The elution buffer contained 200 mM imidazole; 0.5 M NaCl; 20 mM Tris-HCl, at pH 7.9 and 8 M urea. Enzyme concentration of the eluant was around 1.5 mg/ml based on O.D. at 280 nm.

Next, the Hsplase enzyme isolated from the inclusion body, was refolded as follows: the eluant was first adjusted to a final concentration of 4 mM reduced glutathione, 6 M urea and 0.5 mg/ml enzyme. The sample was then dialyzed step-wise against refolding buffers containing 50 mM Tris-Phosphate, pH 7.9, 100 mM NaCl, 1 mM reduced glutathione with urea concentrations of 4 M, 2 M, and 1 M. The refolding buffer was changed every 8-12 hours. Finally, the sample was dialyzed against the refolding buffer with 20% glycerol. After refolding, the sample was passed through a 0.41 μm filter and loaded on to a 4 ml His-bind column equilibrated with the refolding buffer. The column was then washed with the refolding buffer further containing 60 mM imidazole, then eluted with the refolding buffer further containing 500 mM imidazole. The fractions containing enzyme were pooled and desalted on a 70 ml Sephadex G-25 column equilibrated with refolding buffer containing 20% glycerol. The enzyme sample (c.a. 1.5 mg/ml) was stored at -70°C. This HSplase that had been isolated from the inclusion body and refolded, was then evaluated; the Spl binding site was shown to specifically bind to Spl sequences. However, the cleavage site of this HSplase that had been isolated from the inclusion body and refolded, appeared to bind nonspecifically and to non-specifically degrade DNA. That is, DNA without Spl site was degraded nonspecifically. The enzyme was shown to bind to Spl site specifically. However, little specific DNA cleavage was observed. Accordingly, the Hsplase which was isolated from the inclusion body and refolded is less preferred as a recombinant tool.

Preparation of Substrates for Hsplase

To evaluate the Hsplase as a restriction endonuclease, substrates containing Spl sites were constructed.

Preparation of Plasmid pUCSpl Having a Single Spl Site

A synthetic 26 base pair double-stranded DNA fragment having the following base sequence was prepared:

5'-aat tcg ccg ggg cgg ggc ttc tgc ag-3' Seq.ID No. 16

3'-gc ggc ccc gcc ccg aag acg tcc tag-5'

The synthetic DNA fragment, Seq. ID No. 16 contained the Spl site with its flanking sequence from human metallothionein IIa gene. The Spl site is underlined. The DNA fragment Seq. ID No. 16 which has EcoRI/BamHI "sticky ends", was then ligated into the poly linker site of a EcoRI/BamHI-cleaved pUC19 plasmid to produce the plasmid designated "pUCSpl". The plasmid pUCSpl has a single Spl site. Since the synthetic DNA fragment Seq. ID No. 16 lacked 5'-phosphate groups, only one such fragment could be ligated into each pUC19 plasmid. Several clones were picked and a positive clone was identified by analyzing its plasmid DNAs for the absence of KpnI site present between EcoRI and BamHI sites in the polylinker region of pUC19. The sequence of pUCSPl was confirmed by dideoxy-sequencing.

Preparation of Plasmids Having Multiple Spl sites: pUC2Spl, pUC3Spl, pUC4Spl, pUC5Spl and pUC6Spl

Plasmids having multiple Spl sites were constructed by inserting from one to five copies of a 24-mer synthetic DNA fragment Seq. ID No. 17, into pUCSpl at the XbaI site.

The 24-mer synthetic DNA fragment Seq. ID No. 17, which contains a single Spl site and XbaI sticky ends, has the following sequence:

5'-cta ggc cgg ggc ggg gct tct gca-3' Seq.ID No. 17
3'-cg gcc ccg ccc cga aga cgt gat c-5'

The synthetic DNA Seq. ID No. 17 was then mixed with XbaI cleaved pUCSpl.

DNA and ligated into pUCSPl at the XbaI site.

The ligation mixture was transformed into JM105 cell. The number of copy(s) of Spl site DNA inserted was determined by electrophoresis of EcoRI/SphI digested plasmid DNAs on 4% metaphor agarose gel. The resulting plasmids contain two to six copies of Spl site(s) and are termed pUC2SPl, pUC3SPl, pUC4SPl, pUC5SPl and pUC6SPl, respectively.

Evaluation of HSplase Enzyme

Specific binding of HSplase to Spl sequence

The ability of Hsplase enzyme to bind specifically to Spl DNA sequence was demonstrated by band-shift assay. 0.5 μg of the of the Hsplase enzyme that had been isolated from the inclusion body and refolded, was mixed with 0.4 μg 26 base pair synthetic DNA substrate Seq. ID No. 16 in a total volume of 10 μl. The DNA substrate Seq ID No. 16 contained a single Spl site. In a separate control sample, the HSplase was mixed with 0.4 μg DNA substrate Seq. ID No. 16 and 2 μg competitor DNA which lacked an Spl site. The samples were incubated in a solution containing 10 mM Tris-HCl, pH 8.0, 50 mM NaCl, 2 mM DTT 0.1 mM $ZnCl_2$, at 25°C for 20 minutes. The samples were applied to the 4% agarose gel.* The mixture of the DNA substrate Seq. ID No. 16 and Hsplase enzyme was applied to Lane 1. The DNA substrate Seq. ID No. 16 was applied to Lane 2. The DNA substrate Seq. ID No. 16 incubated with Hsplase in the presence of the competitor DNA was applied to lane 3. The DNA substrate Seq. ID No. 16 and competitor DNA was applied to lane 4. The competitor DNA incubated with Hsplase was applied to lane 5. The competitor DNA was applied to lane 6. The gel was subjected to electrophoresis in TAE buffer without EDTA. The gel is shown in Figure 1.

As can be seen by the shifted band in lane 1 of the gel shown in Figure 1, splase enzyme readily formed a protein-DNA complex band when mixed with the substrate DNA Seq. ID No. 16 which contains one Spl site. The shifted band is also present in lane 3 which contained the Hsplase incubated with DNA substrate Seq. ID No. 16 and the competitor DNA which lacks Spl site. However, as seen in lane 5, no Hsplase-DNA complex was formed when the competitor DNA alone was incubated with Hsplase, which demonstrates that the binding of splase to the DNA is specific.

Hsplase Cleavage of Circular DNA

To test the ability of Hsplase to cleave circular DNA, two plasmids, pUC19 which lacks an Spl site, and pUC5SPl, which has 5 Spl sites, were subjected to Hsplase digestion.

The reaction mixtures, which had a total volume of 20 μl, contained the following: 0.5 μg for pUC19 and pUCSpl, and 1 μg for pUC3Spl and pUC5Spl. Tris-HCl; 2 mM $MgCl_2$; 5 mM DTT; 0.1 mM $ZnCl_2$; and 10 ng Hsplase. The reaction mixture was incubated at 37°C for 2 hours and then applied to 0.7% agarose gel. The pUC19 DNA-Hsplase mixture was applied to lane 1. pUC19 DNA without Hsplase was applied to lane 2. DNA molecular weight marker VII from Boehringer Mannheim was applied to lanes 3. The pUCSpl-H-splase mixture was applied to lane 4.

The pUCSpl DNA was applied to lane 5. The pUC3SPl incubated

with Hsplase was applied to lane 6, pUC3Spl was applied to lane 7, pUC5Spl incubated with Hsplase was applied to lane 8. The gel was then subject to electrophoresis, and is shown in Figure 2A.

As can be seen in Figure 2A, the Hsplase cleaved the pUC5Spl DNA near the Spl sites and converted the circular DNA to a linear form which migrated as a 2.7Kb fragment, as shown in lane 8. Little nonspecific nuclease activity is shown in Figure 2A, lane 8.

To confirm that Hsplase cleaved the plasmid DNA specifically near Spl sites, the linearized DNA that is, the DNA bands migrating at position of 2.7 kb. in Figure 2A were excised from lanes 1 & 8 of agarose gel and purified using gene-clean kit from BIO 101, USA. This linearized DNA was then digested by a second restriction enzyme, AlwN I. Specifically, the excised DNAs were incubated with 10 units of AlwNI in buffer supplied by manufacturer at 37°C for 1 hour and applied to a 0.8% agarose gel. pUC19 substrate after double digestion was applied to lane 1. pUC5SPl substrate after double digestion was applied to lane 2. The DNA molecular weight marker VII from Boehringger Mannheim was applied to lane 3. The gel was then subjected to electrophoresis and is shown in Figure 2B.

The linearized plasmid from lane 8 of the gel shown in Figure 2A was cleaved into two fragments, a 1.9 kb fragment and a 0.8 kb fragment. These fragments are shown in Figure 2B, lane 2. This confirms that the cut was near the Spl sites.

In contrast, incubating of the control DNA pUC19 with the Hsplase enzyme yield minor quantity of DNA migrating with a size of 2.7 kb as shown in Figure 2A, lane 1. Subsequent digestion of this band by AlwNI did not yield any new fragment, as shown in Figure 2, B, lane 1. These results indicated that the Hsplase enzyme was specific for Spl site and had little nonspecific nuclease activity.

HSplase Cleavage of Linear DNA

The plasmid, pUC-BENN-CAT which contains the LTR sequence of HIV, along with pUC19 and pUC5spl were used as substrates for Hsplase. The LTR sequence contains the following spl sites:

5'-GAAGCGTGGC-3' Sequence ID No. 7

5'-TGGGCGGGAC-3' Sequence ID No. 8

5'-GGGGAGTGGC-3' Sequence ID No. 9

The plasmid DNAs of pUC19 and pUC5Spl were first linearized by the restriction enzyme AlwNI, while the PUC-Benn-Cat was linearized by digestion with Bam HI. These linearized plasmids were then subject to restriction digestion with Hsplase. The reaction mixtures had a total volume of 10 μl and contained the 10 mM Tris-HCl, 0.75 μg of plasmid DNA, 2 mM $MgCl_2$, 5 mM DTT, 0.1 mM $ZnCl_2$, 100 μg/ml BSA and 10 ng of H-splase. The reaction was incubated at 37°C for 2 or 3 hours and

applied to 0.7% agarose gel. Molecular weight markers were applied to lanes 1 and 8; pUC19 linearized by AlwNI was applied to lane 2; AWNI linearized pUC19 incubated with Hsplase was applied to lane 3; pUC5Spl linearized by AlwNI was applied to lane 4; AlWNI linearized pUC5Spl incubated with Hsplase was applied to lane 5; pUC-BENN-CAT linearized by BamHI was applied to lane 6. The BamHI linearized pUC-BENN-CAT incubated with Hsplase was applied to lane 7. The gels were then subject to electrophoresis.

As shown in Figure 3, the linearized pUC19 control DNA is not cut by Hsplase (lanes 2-3), while pUC5Spl is cut specifically into 1.9 kb and 0.8 kb fragment as shown in lanes 4-5. Most importantly, the linearized pUC-BENN-CAT which carries the HIV LTR sequence with three consecutive Spl sites was also cut specifically by splase. As shown in lanes 6 and 7 of Figure 3, cleavage of the BamHI-linearized pUC-BENN-CAT by splase near Spl sites generated two fragments of 4.20 kb and 1.74 kb.

In addition to the methods used herein to produce the splase enzyme, the enzyme may be made using conventional techniques such as peptide synthesisers for assembling amino acids.

Conclusion

The fusion gene encoding the splase enzyme, a hybrid endonuclease has been constructed and expressed in E. coli. The splase enzyme , purified from E. coli binds specifically to Spl DNA site and digests plasmid DNAs carrying Spl sites. The Hsplase enzyme is also a relatively specific rare-cutter restriction endonuclease. Splase and Hsplase enzymes are both efficient, specific and useful for practical application in biotechnology techniques.

The present invention includes: the DNA sequences encoding a restriction endonuclease comprising the cleavage domain of FokI and the binding domain of Spl, the messenger RNA transcript of such DNA sequence; and the restriction endonuclease which recognizes spl sites.

For example, the DNA sequences include: DNA molecules which, but for the degeneracy of the genetic code would hybridize to DNA encoding the artificial restriction nuclease, thus the degenerate DNA which encodes the artificial restriction nuclease; DNA strands complementary to DNA sequences encoding the artificial restriction nuclease or portions thereof including DNA in SEQ ID 1 and 3 or portions thereof; heterologous DNA having substantial sequence homology to the DNA encoding the artificial restriction nuclease, including the DNA sequences in SEQ ID NO 2 and 4 or portions thereof.

The artificial restriction nuclease includes, for example, artificial restriction endonuclease comprising the portions of cleavage domain of FokI and the binding domain of Spl and proteins having

substantially the same amino acid sequence as shown in SEQ ID NO 2 and 4 or portions thereof.

While the invention has been described with a certain degree of particularity, various adaptations and modifications can be made without departing from the scope of the invention as defined in the appended claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Tsai, Ming-Daw
Huang, Baohua

(ii) TITLE OF INVENTION: Artificial Restriction Endonuclease

(iii) NUMBER OF SEQUENCES: 17

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Calfee, Halter and Griswold
(B) STREET: Suite 1800, 800 Superior Avenue
(C) CITY: Cleveland
(D) STATE: Ohio
(E) COUNTRY: U.S.A.
(F) ZIP: 44114-2688

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Golrick, Mary E.
(B) REGISTRATION NUMBER: 34,829
(C) REFERENCE/DOCKET NUMBER: 18525/00110

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (216) 622-8458
(B) TELEFAX: (216) 241-0816

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 894 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..909

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..894

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
ATG GCT AAA AAG AAA CAG CAT ATT TGC CAC ATC CAA GGC TGT GGG AAA          48
Met Ala Lys Lys Lys Gln His Ile Cys His Ile Gln Gly Cys Gly Lys
  1               5                  10                  15

GTG TAT GGC AAG ACC TCT CAC CTG CGG GCA CAC TTG CGC TGG CAT ACA          96
Val Tyr Gly Lys Thr Ser His Leu Arg Ala His Leu Arg Trp His Thr
             20                  25                  30

GGC GAG AGG CCA TTT ATG TGT ACC TGG TCA TAC TGT GGG AAA CGC TTC         144
Gly Glu Arg Pro Phe Met Cys Thr Trp Ser Tyr Cys Gly Lys Arg Phe
         35                  40                  45

ACA CGT TCG GAT GAG CTA CAG AGG CAC AAA CGT ACA CAC ACA GGT GAG         192
Thr Arg Ser Asp Glu Leu Gln Arg His Lys Arg Thr His Thr Gly Glu
     50                  55                  60

AAG AAA TTT GCC TGC CCT GAG TGT CCT AAG CGC TTC ATG AGG AGT GAC         240
Lys Lys Phe Ala Cys Pro Glu Cys Pro Lys Arg Phe Met Arg Ser Asp
 65                  70                  75                  80

CAC CTG TCA AAA CAT ATC AAG ACC CAC CAG AAT AAG AAG GTA CCT AAT         288
His Leu Ser Lys His Ile Lys Thr His Gln Asn Lys Lys Val Pro Asn
                 85                  90                  95

CGT GGT GTG ACT AAG CAA CTA GTC AAA AGT GAA CTG GAG GAG AAG AAA         336
Arg Gly Val Thr Lys Gln Leu Val Lys Ser Glu Leu Glu Glu Lys Lys
            100                 105                 110

TCT GAA CTT CGT CAT AAA TTG AAA TAT GTG CCT CAT GAA TAT ATT GAA         384
Ser Glu Leu Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr Ile Glu
        115                 120                 125

TTA ATT GAA ATT GCC AGA AAT TCC ACT CAG GAT AGA ATT CTT GAA ATG         432
Leu Ile Glu Ile Ala Arg Asn Ser Thr Gln Asp Arg Ile Leu Glu Met
    130                 135                 140

AAG GTA ATG GAA TTT TTT ATG AAA GTT TAT GGA TAT AGA GGT AAA CAT         480
Lys Val Met Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly Lys His
145                 150                 155                 160

TTG GGT GGA TCA AGG AAA CCG GAC GGA GCA ATT TAT ACT GTC GGA TCT         528
Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser
                165                 170                 175

CCT ATT GAT TAC GGT GTG ATC GTG GAT ACT AAA GCT TAT AGC GGA GGT         576
Pro Ile Asp Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly
            180                 185                 190

TAT AAT CTG CCA ATT GGC CAA GCA GAT GAA ATG CAA CGA TAT GTC GAA         624
Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr Val Glu
        195                 200                 205

GAA AAT CAA ACA CGA AAC AAA CAT ATC AAC CCT AAT GAA TGG TGG AAA         672
Glu Asn Gln Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp Trp Lys
    210                 215                 220

GTC TAT CCA TCT TCT GTA ACG GAA TTT AAG TTT TTA TTT GTG AGT GGT         720
Val Tyr Pro Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser Gly
225                 230                 235                 240

CAC TTT AAA GGA AAC TAC AAA GCT CAG CTT ACA CGA TTA AAT CAT ATC         768
His Phe Lys Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn His Ile
                245                 250                 255
```

```
ACT AAT TGT AAT GGA GCT GTT CTT AGT GTA GAA GAG CTT TTA ATT GGT        816
Thr Asn Cys Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly
            260                 265                 270

GGA GAA ATG ATT AAA GCC GGC ACA TTA ACC TTA GAG GAA GTG AGA CGG        864
Gly Glu Met Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg
        275                 280                 285

AAA TTT AAT AAC GGC GAG ATA AAC TTT TAA                                894
Lys Phe Asn Asn Gly Glu Ile Asn Phe
    290                 295
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 297 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Met Ala Lys Lys Lys Gln His Ile Cys His Ile Gln Gly Cys Gly Lys
 1               5                  10                  15

Val Tyr Gly Lys Thr Ser His Leu Arg Ala His Leu Arg Trp His Thr
            20                  25                  30

Gly Glu Arg Pro Phe Met Cys Thr Trp Ser Tyr Cys Gly Lys Arg Phe
        35                  40                  45

Thr Arg Ser Asp Glu Leu Gln Arg His Lys Arg Thr His Thr Gly Glu
    50                  55                  60

Lys Lys Phe Ala Cys Pro Glu Cys Pro Lys Arg Phe Met Arg Ser Asp
65                  70                  75                  80

His Leu Ser Lys His Ile Lys Thr His Gln Asn Lys Lys Val Pro Asn
                85                  90                  95

Arg Gly Val Thr Lys Gln Leu Val Lys Ser Glu Leu Glu Glu Lys Lys
            100                 105                 110

Ser Glu Leu Arg His Lys Leu Lys Tyr Val Pro His Glu Tyr Ile Glu
        115                 120                 125

Leu Ile Glu Ile Ala Arg Asn Ser Thr Gln Asp Arg Ile Leu Glu Met
    130                 135                 140

Lys Val Met Glu Phe Phe Met Lys Val Tyr Gly Tyr Arg Gly Lys His
145                 150                 155                 160

Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser
                165                 170                 175

Pro Ile Asp Tyr Gly Val Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly
            180                 185                 190

Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu Met Gln Arg Tyr Val Glu
        195                 200                 205

Glu Asn Gln Thr Arg Asn Lys His Ile Asn Pro Asn Glu Trp Trp Lys
    210                 215                 220
```

```
Val Tyr Pro Ser Ser Val Thr Glu Phe Lys Phe Leu Phe Val Ser Gly
225                 230                 235                 240

His Phe Lys Gly Asn Tyr Lys Ala Gln Leu Thr Arg Leu Asn His Ile
                245                 250                 255

Thr Asn Cys Asn Gly Ala Val Leu Ser Val Glu Glu Leu Leu Ile Gly
            260                 265                 270

Gly Glu Met Ile Lys Ala Gly Thr Leu Thr Leu Glu Glu Val Arg Arg
        275                 280                 285

Lys Phe Asn Asn Gly Glu Ile Asn Phe
    290                 295
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 912 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..909

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATG GCT CAT CAC CAT CAC CAT CAC AAA AAG AAA CAG CAT ATT TGC CAC        48
Met Ala His His His His His His Lys Lys Lys Gln His Ile Cys His
  1               5                  10                  15

ATC CAA GGC TGT GGG AAA GTG TAT GGC AAG ACC TCT CAC CTG CGG GCA        96
Ile Gln Gly Cys Gly Lys Val Tyr Gly Lys Thr Ser His Leu Arg Ala
            20                  25                  30

CAC TTG CGC TGG CAT ACA GGC GAG AGG CCA TTT ATG TGT ACC TGG TCA       144
His Leu Arg Trp His Thr Gly Glu Arg Pro Phe Met Cys Thr Trp Ser
        35                  40                  45

TAC TGT GGG AAA CGC TTC ACA CGT TCG GAT GAG CTA CAG AGG CAC AAA       192
Tyr Cys Gly Lys Arg Phe Thr Arg Ser Asp Glu Leu Gln Arg His Lys
    50                  55                  60

CGT ACA CAC ACA GGT GAG AAG AAA TTT GCC TGC CCT GAG TGT CCT AAG       240
Arg Thr His Thr Gly Glu Lys Lys Phe Ala Cys Pro Glu Cys Pro Lys
65                  70                  75                  80

CGC TTC ATG AGG AGT GAC CAC CTG TCA AAA CAT ATC AAG ACC CAC CAG       288
Arg Phe Met Arg Ser Asp His Leu Ser Lys His Ile Lys Thr His Gln
                85                  90                  95

AAT AAG AAG GTA CCT AAT CGT GGT GTG ACT AAG CAA CTA GTC AAA AGT       336
Asn Lys Lys Val Pro Asn Arg Gly Val Thr Lys Gln Leu Val Lys Ser
            100                 105                 110
```

```
GAA CTG GAG GAG AAG AAA TCT GAA CTT CGT CAT AAA TTG AAA TAT GTG        384
Glu Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys Leu Lys Tyr Val
        115                 120                 125

CCT CAT GAA TAT ATT GAA TTA ATT GAA ATT GCC AGA AAT TCC ACT CAG        432
Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg Asn Ser Thr Gln
    130                 135                 140

GAT AGA ATT CTT GAA ATG AAG GTA ATG GAA TTT TTT ATG AAA GTT TAT        480
Asp Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe Met Lys Val Tyr
145                 150                 155                 160

GGA TAT AGA GGT AAA CAT TTG GGT GGA TCA AGG AAA CCG GAC GGA GCA        528
Gly Tyr Arg Gly Lys His Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala
                165                 170                 175

ATT TAT ACT GTC GGA TCT CCT ATT GAT TAC GGT GTG ATC GTG GAT ACT        576
Ile Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val Ile Val Asp Thr
            180                 185                 190

AAA GCT TAT AGC GGA GGT TAT AAT CTG CCA ATT GGC CAA GCA GAT GAA        624
Lys Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu
        195                 200                 205

ATG CAA CGA TAT GTC GAA GAA AAT CAA ACA CGA AAC AAA CAT ATC AAC        672
Met Gln Arg Tyr Val Glu Glu Asn Gln Thr Arg Asn Lys His Ile Asn
    210                 215                 220

CCT AAT GAA TGG TGG AAA GTC TAT CCA TCT TCT GTA ACG GAA TTT AAG        720
Pro Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val Thr Glu Phe Lys
225                 230                 235                 240

TTT TTA TTT GTG AGT GGT CAC TTT AAA GGA AAC TAC AAA GCT CAG CTT        768
Phe Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr Lys Ala Gln Leu
                245                 250                 255

ACA CGA TTA AAT CAT ATC ACT AAT TGT AAT GGA GCT GTT CTT AGT GTA        816
Thr Arg Leu Asn His Ile Thr Asn Cys Asn Gly Ala Val Leu Ser Val
            260                 265                 270

GAA GAG CTT TTA ATT GGT GGA GAA ATG ATT AAA GCC GGC ACA TTA ACC        864
Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly Thr Leu Thr
        275                 280                 285

TTA GAG GAA GTG AGA CGG AAA TTT AAT AAC GGC GAG ATA AAC TTT            909
Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe
    290                 295                 300

TAA                                                                    912
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 303 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Ala His His His His His His Lys Lys Lys Gln His Ile Cys His
  1               5                  10                  15

Ile Gln Gly Cys Gly Lys Val Tyr Gly Lys Thr Ser His Leu Arg Ala
```

```
            20                  25                  30
His Leu Arg Trp His Thr Gly Glu Arg Pro Phe Met Cys Thr Trp Ser
        35                  40                  45
Tyr Cys Gly Lys Arg Phe Thr Arg Ser Asp Glu Leu Gln Arg His Lys
    50                  55                  60
Arg Thr His Thr Gly Glu Lys Lys Phe Ala Cys Pro Glu Cys Pro Lys
65                  70                  75                  80
Arg Phe Met Arg Ser Asp His Leu Ser Lys His Ile Lys Thr His Gln
                85                  90                  95
Asn Lys Lys Val Pro Asn Arg Gly Val Thr Lys Gln Leu Val Lys Ser
            100                 105                 110
Glu Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys Leu Lys Tyr Val
        115                 120                 125
Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg Asn Ser Thr Gln
    130                 135                 140
Asp Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe Met Lys Val Tyr
145                 150                 155                 160
Gly Tyr Arg Gly Lys His Leu Gly Gly Ser Arg Lys Pro Asp Gly Ala
                165                 170                 175
Ile Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val Ile Val Asp Thr
            180                 185                 190
Lys Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly Gln Ala Asp Glu
        195                 200                 205
Met Gln Arg Tyr Val Glu Glu Asn Gln Thr Arg Asn Lys His Ile Asn
    210                 215                 220
Pro Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val Thr Glu Phe Lys
225                 230                 235                 240
Phe Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr Lys Ala Gln Leu
                245                 250                 255
Thr Arg Leu Asn His Ile Thr Asn Cys Asn Gly Ala Val Leu Ser Val
            260                 265                 270
Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala Gly Thr Leu Thr
        275                 280                 285
Leu Glu Glu Val Arg Arg Lys Phe Asn Asn Gly Glu Ile Asn Phe
    290                 295                 300
```

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

KGGGCGGRRY 10

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

GGGGCGGGGC 10

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GAAGCGTGGC 10

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TGGGCGGGAC 10

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GGGGAGTGGC 10

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

KRRGMGKRRY 10

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 34 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

GTCCATGGCT AAAAAGAAAC AGCATATTTG CCAC 34

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CCGGTACCTT CTTATTCTGG TGGGTC

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

CGGGTACCTA ATCGTGGTGT GACTAAG 27

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GCGGATCCTT AAAAGTTTAT CTCGCCGTTA TT

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 52 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

GTCCATGGCT CATCACCATC ACCATCACAA AAAGAAACAG CATATTTGCC AC 52

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 26 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

AATTCGCCGG GGCGGGGCTT CTGCAG 26

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CTAGGCCGGG GCGGGGCTTC TGCA 24

WHAT IS CLAIMED:

1. An artificial restriction endonuclease comprising the cleavage domain of FokI and the binding domain of Spl.

2. The endonuclease of claim 1 further comprising from 6 to 10 histidines at the amino terminus.

3. The endonuclease of claim 1 wherein the endonuclease comprises the amino acid sequence as shown in Sequence ID No. 2.

4. The endonuclease of claim 1 wherein the endonuclease comprises the amino acid sequence as shown in Sequence ID No. 4.

5. An artificial gene encoding artificial restriction endonuclease of claim 1.

6. The gene of claim 3 wherein the gene codes for the enzyme as shown in Sequence ID No. 2.

7. The gene of claim 3 wherein the gene codes for the enzyme as shown in Sequence ID No. 4.

8. The gene of claim 4 wherein the gene codes for the endonuclease of claim 2.

9. The artificial restriction endonuclease of claim 1, wherein the endonuclease recognizes the following nucleotide sequence: 5'-G(T)G(A)G(A) GC(A)G G(T)G(A)G(A)C(T)-3'.

10. An artificial restriction endonuclease consisting essentially of the cleavage domain of FokI and the binding domain of Spl.

11. A vector containing the gene of claim 5.

12. A cell containing the gene of claim 5.

13. A method for cleaving both circular and linear DNA samples having at least one Spl site comprising the following steps:

(a) providing the artificial restriction endonuclease of claim 1;

(b) mixing said artificial restriction endonuclease with the DNA sample;

(c) incubating the DNA sample with the artificail restriction endonuclease for a time sufficient to provide one or more DNA fragments where the DNA was circular, or to provide at least two DNA fragments where the DNA was linear.

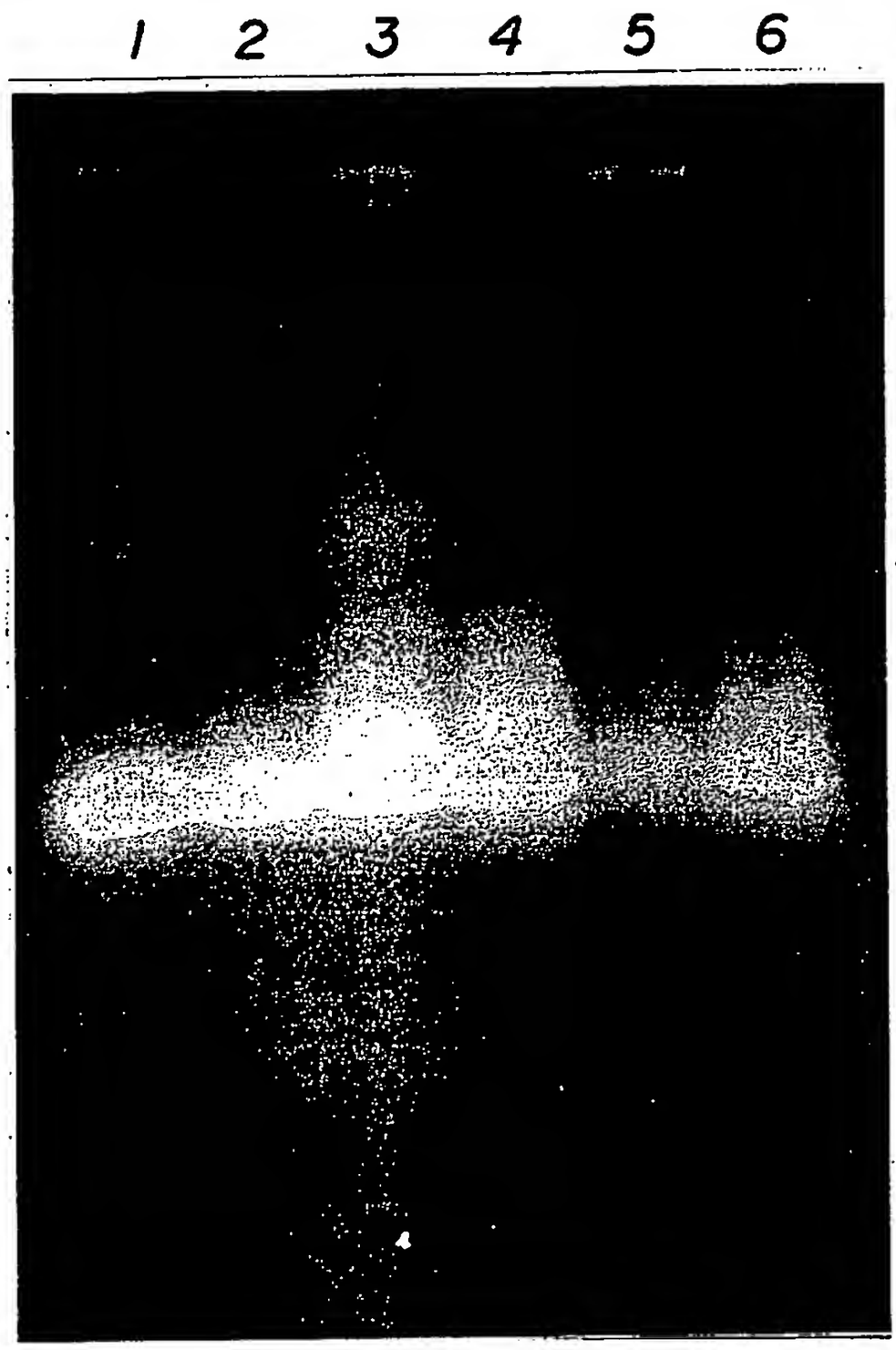

FIG. 1

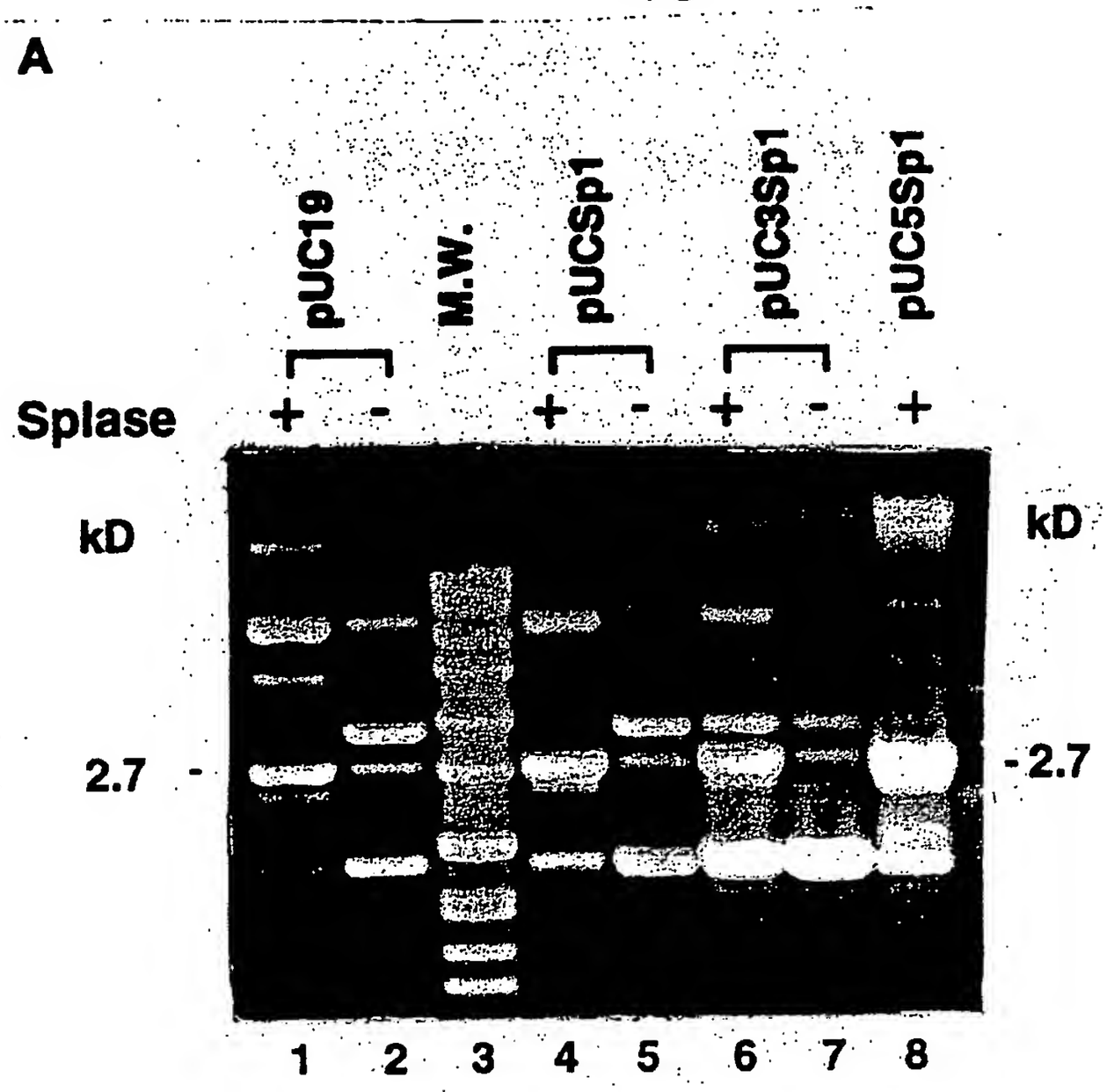
A
pUC19
M.W.
pUCSp1
pUC3Sp1
pUC5Sp1
Splase
+ - + - + - +
kD
kD
2.7
2.7
1 2 3 4 5 6 7 8

FIG. 2A

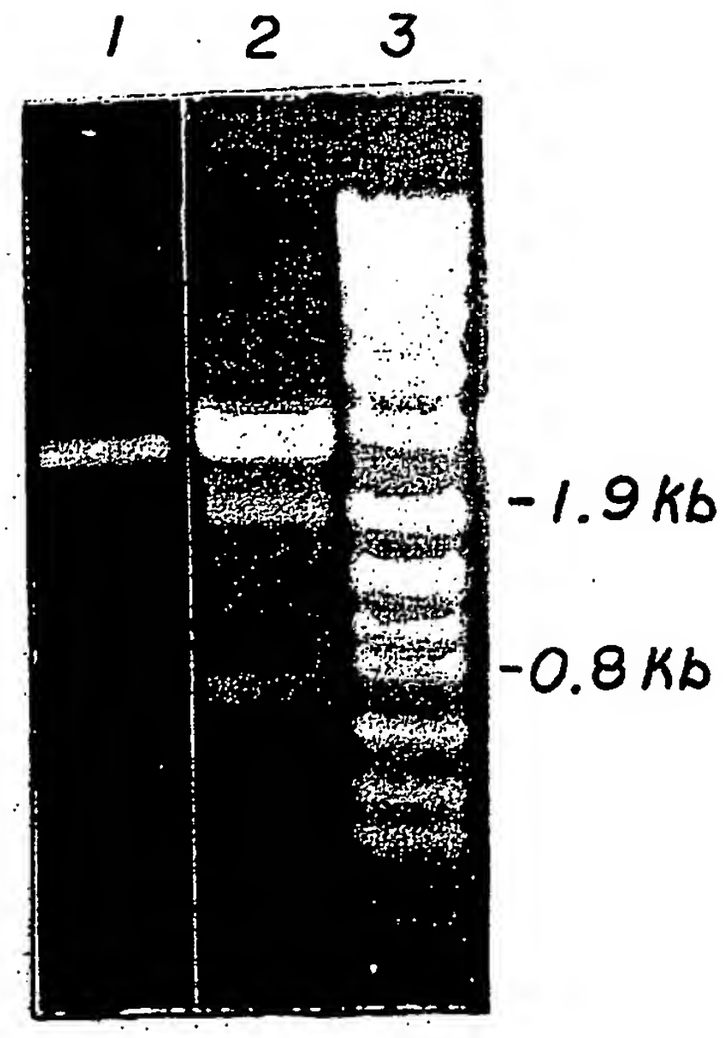
1 2 3
1.9 Kb
0.8 Kb

FIG. 2B

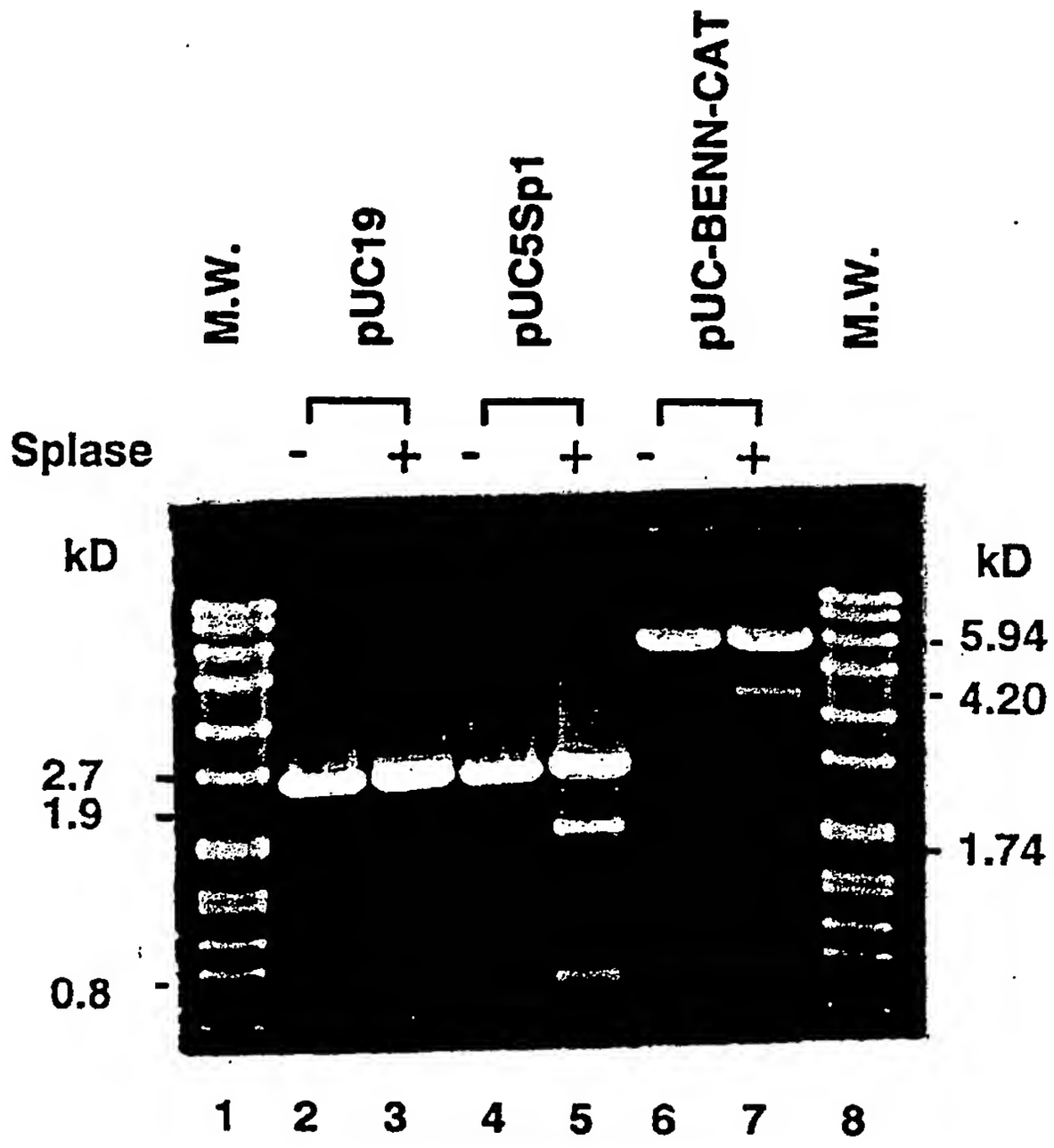
M.W.
pUC19
pUC5Sp1
pUC-BENN-CAT
M.W.
Splase
- + - + - +
kD
kD
5.94
4.20
2.7
1.9
1.74
0.8
1 2 3 4 5 6 7 8

FIG. 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/09315

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :C12N 9/22, 15/55; C12P 19/34
US CL :435/91.1, 199; 536/23.2

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/91.1, 199; 536/23.2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5,356,802 A (CHANDRASEGARAN) 18 October 1994 (18.10.94), column 5, line 25 - column 6, line 25.	1-13
Y, P	US 5,436,150 A (CHANDRASEGARAN) 25 July 1995 (25.07.95), column 8, line 27 - column 9, line 53 and column 22, line 1 - column 24, line 43.	1-13
A, P	US 5,487,994 A (CHANDRASEGARAN) 30 January 1996 (30.01.96).	1-13

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
07 AUGUST 1996	26 AUG 1996

Name and mailing address of the ISA/US	Authorized officer
Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231	CHARLES PATTERSON
Facsimile No. (703) 305-3230	Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)★